Classification of African ecosystems at 1 km resolution using multiannual SPOT/VEGETATION data and a hybrid clustering approach

Armel T. Kaptué Tchuenté, Steven de Jong, Jean-Louis Roujean, Charly Favier, Catherine Mering

Classification of African ecosystems at 1 km resolution using multiannual SPOT/VEGETATION data and a hybrid clustering approach

Armel T. Kaptué Tchuenté[1], Steven M. De Jong[2], Jean-Louis Roujean[1], Charly Favier[3], Cathérine Mering[4]

1. *Centre National de Recherches Météorologiques, Toulouse, France*
2. *Physical Geography Research Institute, Utrecht, The Netherlands*
3. *Institut des Sciences de l'Evolution, Montpellier, France*
4. *Pôle de Recherche pour l'Organisation et la Diffusion de l'Information Géographique, Paris, France*

Email: kaptue@cnrm.meteo.fr

ABSTRACT

Ecosystems classification is the process of allocating vegetation types into groups so that individuals in the same class are similar according to their physiological and phonological characteristics to another one. Over large areas, the only suitable technique to obtain frequent and repetitive data acquisitions over such large areas is the use of observations recorded by sensors of moderate resolution. In order to minimize the role of the analyst and to improve the accuracy of the results, innovative and efficient approaches for the classification of ecosystems continue to appear in the literature. This research developed and implemented a new hybrid unsupervised classification approach to derive ecosystems using multi-annual time series by combining hierarchical and partitioning clustering principles. The latter approach is applied on 8-years time series (2000-2007) of 10-day composite Normalized Difference Vegetation Index (NDVI) recorded by SPOT/VEGETATION. After the first segmentation of the mainland in ecoregions using the Fast Fourier Transform (FFT), successive k-nearest neighbor (k-NN) clustering enhance the discrimination of ecosystems and yields to the production of a new ecosystem map for the African continent. The nomenclature relied on the Land Cover Classification System (LCCS) of the Food and Agricultural Organization (FAO). On the basis of validated continental, a pixel-by-pixel analysis is conducted to assess the accuracy of the new classification. The hybrid clustering facilitates the identification/labeling process and the obtained results which should provide key information needed for management/monitoring of natural resources, biodiversity conservation and biogeochemical studies may also deserve vegetation cover modeling at regional and local scale

1 INTRODUCTION

Identification maps of ecosystems foster sustainable strategies for a wide range of human activities (Cihlar, 2000). Ecosystems mirror the ability of biological systems to support human needs such as food security, and provide critical environmental indicators for biotic diversity worldwide. Ecosystems maps produced from compilation of pre-existing atlases and ancillary data analysis are not effective over large areas, as they are very time consuming and date lagged. Remote sensing offer a synoptic view.

Very recently, an ecosystem map based on MODIS time series analysis emphasized recent major changes in Western Africa (Kaptué et al., 2010 a). Other initiatives for this region are AFRICOVER (FAO, 2004), GLC2000 (Mayaux et al., 2004), GLOBCOVER (Bicheron et al., 2008) and the annual MODIS land cover product (Friedl et al., 2010). Up to now, such maps have relied on an analysis of vegetation index that combine spectral reflectances in the red and near infrared shortwaves.

An important aspect in ecosystem classification over large areas is the incorporation of environmental information due to the fact that the heterogeneity of the landscape may result in large variations of the reflectance for the same ecosystem class. A common approach is to stratify the area of interest into zones showing similar spectral behaviour. But this is generally done more or less arbitrarily rather than using a formalism.

The objective of this study is to present an innovative hybrid clustering approach, the so-called *stacked segment classification*, which includes both partitioning and hierarchical clustering principles. The methodology is applied on multiannual VGT NDVI datasets centered over the entire African continent to produce a new ecosystem map. The method is expected to be able to better discriminate vegetation types than existing pan-continental maps which focus only on one year of satellite datasets.

2 DATA POST-PROCESSING

The VEGETATION (VGT) onboard SPOT includes the red and near infrared spectral reflectance bands that are specially design to monitor the phenology of green vegetation and its orbit ensures daily global coverage of the Earth's surface with a 1-km footprint. The datasets used to produce the new ecosystem map consist of 8 years (from January 1, 2000 to December 31, 2007) of VEGETATION (VGT) data. A 4-degree polynomial fit was finally applied in order to fill gaps and discard outliers (Kaptué et al., 2010 a). Human-made surfaces and flooded saltpans were separated by reference to the GLC2000 (Mayaux et al., 2004) and masked during the classification process.

3 METHOD

The post-processed images were used in a classification, which was then evaluated (figure 1-a). The classification process included two unsupervised techniques: the k-NN and the *stacked segment classification*.

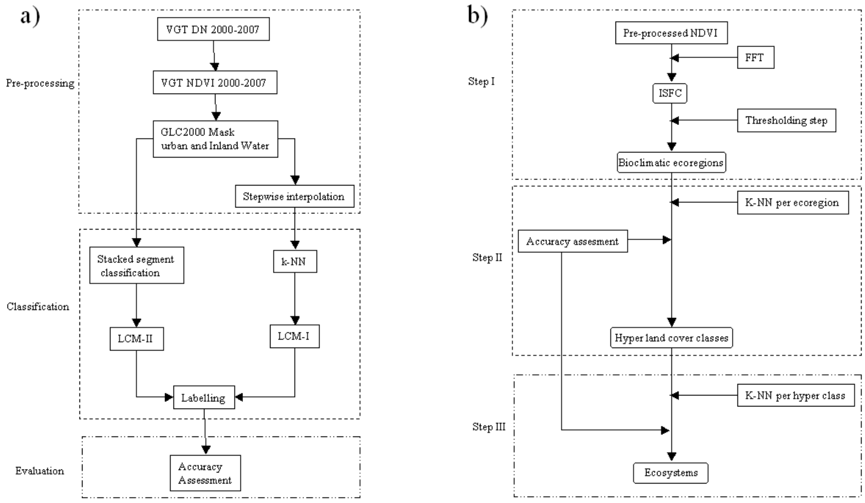

Figure 1: Methodological scheme: a) Flowchart of the project, b) Stacked segment classification

Since Africa extends on either side of the equator, inversions of the seasonal cycle were observed and also high variation values of NDVI within the same. In order to preserve the natural limits of the vegetation, our approach qualifies as a s*tacked segments classification*, is paved by hierarchical and dynamical clustering. It exploits the Fast Fourier Transform (FFT) and the k-NN classifier to produce the second ecosystem map with 73 ecosystems referred to as LCM-II. The mainland has been divided in 5 bioclimatic ecoregions by exploiting the similarities obtained with the Index of Segmentation of Fourier Component (ISFC), a new quantitative measure based on Fourier analysis of NDVI datasets (Kaptué et al., 2010 b).

The objectives of this first stratification of the continent into ecoregions were twofold: (i) to reduce the NDVI variability to improve classification accuracy and, (ii) to enhance regional trends like seasonality, cloud coverage, etc. Using this stratification, a k-NN clustering was performed for each region on the basis of the temporal information of the NDVI time series. This yielded 18 subsets. Finally, a k-NN clustering was performed separately on each of the 18 subsets to produce 192 clusters. Vegetation types encountered in the same bioclimatic ecoregion were merged and labelled to make up the second ecosystem map (LCM-II hereafter). The different steps of this hybrid approach are summarised in figure 1-b.

4 RESULTS

4.1 Bioclimatic ecoregions

Figure 2-a shows the ISFC components from an FFT-like transformation applied to the multi-annual time series of satellite data at the resolution of 1 km for the whole continent. It is marked by a zoning pattern, as expected, with the lowest values lying between 20°N and 30°N. The highest values are found in the equatorial region, between 10° N and 10° S. North of 10°S, ISFC values show a general decline from west to east whereas, south of 10°S, the situation is reversed. In fact, ISFC values tend to be damped in a westward direction. By fixing a threshold on ISFC values somewhat arbitrarily, the continent can be split into 5 equal bioclimatic ecoregions (figure 2-b). In addition, the distance to the equator and the coastlines influences local variability.

4.2 Ecosystem units

African forests include dry tropical forest in both eastern and southern Africa, then humid tropical forest in western and central parts of the continent. Each of these types is characterized by abiotic conditions such as the vegetation structure (stand height, vertical stratification) and specific floristic composition. For instance, closed evergreen forests showing a density of more than 70% for an upper stratum 35-45 metres high precludes the development of shrub and grass strata. At the geographic boundaries of forested areas, the transition to arid and semi-arid domains is marked by a sparse distribution of woodlands and shrublands with possible association with grassland. Such a herbaceous stratum characterizes regions having seasonally waterlogged soil. It encompasses mainly savannahs and fallows, which can be mixed with cultures due to the low intensification level of agricultural techniques. This induces inconsistencies over time and across countries among the many sources of agricultural inventory data. Nevertheless, large plantations are characterized by a uniform cover and follow a geometrical layout.

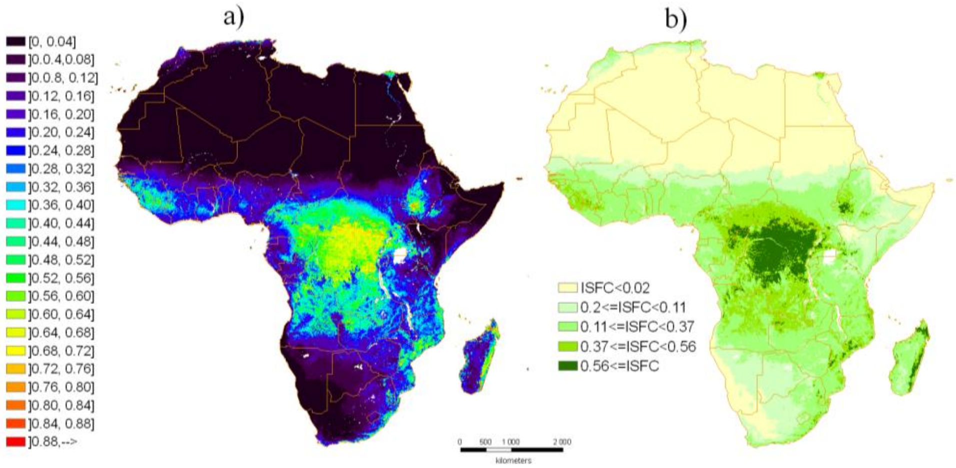

Figure 2: a) The Index of Segmentation of Fourier Component; b) The bioclimatic ecoregions

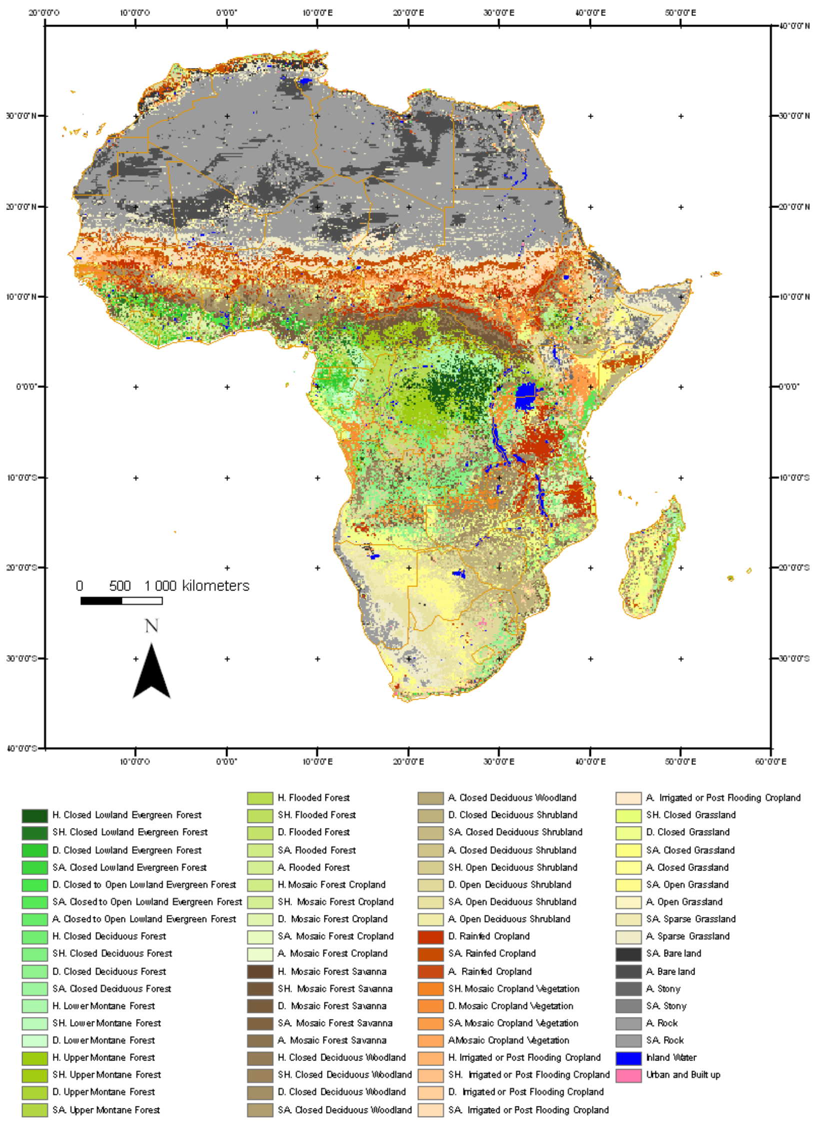

Figure 3: The second ecosystem map

Two map products were made at continental scale. One came from the single k-NN classifier while a second was based on the *stacked segment classification* approach (figure 1). Figures 3 provide the location and extent of ecosystems types in Africa for LCM-II. In this map, some classes present a particular physiognomy at a given place or at a given time period. For instance, *Irrigated croplands* are mainly found along Nile valley. We note a category of *open grassland* (semi-arid) mainly found in South Africa. The *deciduous forest* is located on some humid mountainsides and ravines of Chaillu or Mayombe (north of Congo). As for the *closed lowland forest*, we delineate four facies corresponding to a humid closed semi-evergreen formation from the Atlantic coast to the Sudano-Zambezian savannah in the southwest of the Central African Republic or southeast Cameroon. This permits the isolation of Gabon Forest included in the the coastal facies of the *closed lowland forest* tothe Mayombe transborder forest among Angola, DRC and Republic of Congo and, finally, other internal forests in DRC. The Batéké region is surrounded by closed lowland forest and formed by grassland along the Congo border. The classify map LCM-II was validated though a comparison with three continental maps: the 1km-GLC2000 map for the year 2000, the 300m-GLOBCOVER map for the year 2005 and the first layer of the 500m MODIS land cover product for the year 2005 (denoted MODIS LC-I hereafter). When compared with GLC2000, GLOBCOVER and MODIS LC-I, respectively LCM-II yields kappas of 0.505, 0.430 and 0.420, with respect to the same order of map products.

5 CONCLUSION

Ecosystem maps provide key environmental information needed for the management and monitoring of natural resources, the conservation of biodiversity and our understanding of the environment. The growing need for ecosystem information has driven developments in sensing technologies using repetitive coverage at short time intervals and simultaneous measurements over large areas. This study has demonstrated that including phenological cycles of ecosystems in classification procedures is successful.

The ISFC, a new FFT-based index of segmentation for splitting areas into equal-reasoning regions based on bioclimatic, ecological and satellite observation considerations is presented. This index could be used to perform an analysis at any scale (local, national, regional, continental, or global) provided that the required multi-annual data sets exist. The results presented in this paper show the potential of the ISFC to produce a 1-km trimmed bioclimatic ecoregion map conforming to the LCCS nomenclature. This improves the capabilities of the k-NN classifier to characterize the agro-ecological zones of a continent. The application of a *stacked segment classification*, a new hybrid multistage clustering approach that captures gradients in terms of timing, duration, and intensity of the photosynthetic activity in the same ecosystem unit permitted us to distinguish the various ecosystem units.

ACKNOWLEDGEMENTS

The authors are indebted to the VITO centre (www.vgt.vito.be) for putting customized NDVI data sets at our disposal and to the people working on the AFRICOVER database (http://www.africover.org) who generated the national maps of DRC, Egypt and Sudan. We express our gratitude to Elna Van Den Berg of the Agricultural Research Council (South Africa), Justin Moat of the Royal Botanic Gardens Kew (UK) and Freddy Nachtergaele of the Food and Agricultural Organization (Italy) for use of their field data. Pan-continental fields of ecosystem and bioclimatic ecoregion maps can be requested through electronic mail to the authors.

REFERENCES

Bicheron, P. & co-authors (2008) GLOBCOVER: product description and validation report, See http://postel.mediasfrance.org/en/DOWNLOAD/Documents/#globcover (Last accessed May 2009).

Cihlar, J. (2000) Land cover mapping of large areas from satellites: status and research priorities, *Int. J. Rem. Sens.*, **21**, 1093-1114.

FAO (2004) The Africover Initiative, Food and Agricultural Organisation, See www.africover.org/africover_initiative.htm (Last accessed December 2009).

Friedl, M. A., Sulla-Menashe, D, Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010) MODIS Collection 5 global land cover: algorithms refinements and characterization of new datasets, *Rem. Sens. Env.* **114,** 168-182.

Kaptué, T. A, T., Roujean, J.-L. & Faroux, S. (2010 a) ECOCLIMAP-II: an ecosystem classification and land surface parameter database of Western Africa at 1 km resolution for the Africa Monsoon Multidisciplinary (AMMA) project , *Remote Sensing of Environment*, **doi:10.1016/j.rse.2009.12.008.**

Kaptué, T. A. T., De Jong, S., Roujean, J.-L, Fabier, C & Mering, C. (2010 b) Ecosystem mapping at the African continent scale using a hybrid clustering approach based on 1-km resolution multi-annual data from SPOT/VEGETATION, *J. Biogeogr.*, **submitted.**

Mayaux, P., Bartholomé, E., Fritz S. & Belward, A. (2004) A new land cover map of Africa for the year 2000, *J. Biogeogr.* **31**, 861-877.